

Claims

- Sub A17
1. A document retrieval system comprising a user interface and processing means, wherein the user interface is configured to allow a user to enter a query phrase indicative of a subject of interest, and the processing means is operative to select query keywords from the query phrase and allocate positional weightings to the query keywords dependent upon the relative positions of the query keywords within the query phrase.
  2. A document retrieval system according to claim 1, wherein the positional weighting applied to query keywords increases progressively from a low weighting at the beginning of the query phrase to a higher weighting at the end of the query phrase.
  3. A document retrieval system according to claim 2, wherein the positional weighting increases in a substantially linear manner.
  4. A document retrieval system according to any of claims 1 to 3, wherein the positional weightings applied to the query keywords are scaled.
  5. A document retrieval system according to claim 4, wherein the scaling is such that the maximum query keyword positional weighting is one.
  6. A document retrieval system according to any preceding claim, wherein the system is arranged to compare the query phrase with a set of document signature phrases, each document signature phrase being indicative of the contents of a document.
  7. A document retrieval system according to claim 6, wherein each document signature phrase comprises document keywords having positional weightings dependent upon their relative positions within the document signature phrase.
  8. A document retrieval system according to claim 7, wherein comparison of the query phrase and the document signature phrase comprises multiplying the positional

10070810-031102

SubAI 7  
weighting of each query keyword by the positional weighting of a corresponding document keyword.

9. A document retrieval system according to claim 8, wherein the results of the multiplication are added together to provide a sum that is a measure of the relevance of the document represented by the document signature phrase.
10. A document retrieval system according to any preceding claim, wherein in addition to the positional weighting given to query keywords, the query keywords are given relevance weightings dependent upon the perceived relevance of the query keywords to the subject of interest.
11. A document retrieval system according to any preceding claim, wherein a subject of interest to the user is represented within the processing means as an interest phrase comprising interest keywords having positional weightings dependent upon the relative positions of the interest keywords within the interest phrase.
12. A document retrieval system according to claim 11, wherein when the user enters a query phrase, the processing means is arranged to locate an existing interest phrase that satisfies a predetermined degree of correspondence between the query keywords and the interest keywords.
13. A document retrieval system according to claim 12, wherein the user interface allows the user to select words from the returned interest phrase, and add them to the query phrase.
14. A document retrieval system according to claim 12 or claim 13, wherein if more than one interest phrase is returned, the phrases are ordered for the user's review in accordance with the degree of correspondence between the query phrase and the interest phrases.
15. A document retrieval system according to any of claims 12 to 14, wherein the existing interest phrases include interest phrases representative of subjects of interest to other users.

10070810 034100  
20110701

SubA17

16. A document retrieval system according to any of claims 12 to 15, wherein when the system is not being used by a given user, the system augments that user's interest phrases by comparing an interest phrase of the given user with interest phrases of other users, and if an interest phrase of another user is sufficiently similar, providing a copy of that interest phrase for the given user.

17. A document retrieval system according to claim 16, wherein contact information regarding the other user is copied to the given user.

18. A document retrieval system according to claim 16 or claim 17, wherein links to documents found by the other user are provided for the given user.

19. A document retrieval system according to any preceding claim, wherein documents retrieved by the system are selected by the user on the basis of their perceived relevance, and document keywords representative of the selected documents are used to update an interest phrase indicative of an interest of the user.

20. A document retrieval system according to claim 19, wherein the interest phrase is updated by adjusting relevance weightings allocated to interest keywords of the interest phrase.

21. A document retrieval system according to claim 19 or claim 20, wherein the interest phrase is updated by adding keywords to the interest phrase.

22. A document retrieval system according to any of claims 19 to 21, wherein the document keywords are used to create a new interest phrase if they are determined not to be relevant to existing interest phrases.

23. A document retrieval system according to claim 22, wherein the user is requested by the user interface to provide a name for the new interest phrase.

24. A method of summarising the content of a document, the method comprising segmenting the document into sentences, selecting document keywords from the

20250310 031100

SubA17  
sentences, and allocating positional weightings to the document keywords dependant upon the relative positions of the document keywords within the sentence.

25. A method according to claim 24, wherein the positional weighting applied to document keywords increases progressively from a low weighting at the beginning of a sentence to a higher weighting at the end of the sentence.
26. A method according to claim 25, wherein the positional weighting increases in a substantially linear manner.
27. A method according to claim 26, wherein the positional weightings applied to document keywords are scaled.
28. A method according to any of claims 24 to 27, wherein where a document keyword occurs more than once in a sentence, the positional weighting is determined on the basis of an average location of the document keyword within the sentence.
29. A method according to any of claims 24 to 28, wherein similar sentences contained in a document are grouped together, and the largest group is taken to be an indication of the average content of the document.
30. A method according to claim 29, wherein a document signature phrase is generated by combining document keywords from each sentence of the group.
31. A method according to claim 30, wherein each document keyword within the document signature phrase is given a relevance weighting dependent upon the number of times it occurs in the group of sentences.
32. A method according to claim 31, wherein the relevance weighting is increased for those document keywords which are capitalised.
33. A document retrieval system substantially as hereinbefore described with reference to the accompanying figures.

10070810.031102  
"OFFICE"

34. A method of summarising the content of a document substantially as hereinbefore described.

20110810.031102